14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

EVALUATION OF A BLIND METHOD FOR THE ESTIMATION OF HURST'S EXPONENT IN TIME SERIES

Federico Esposti and Maria Gabriella Signorini

Dipartimento di Bioingegneria, Politecnico di Milano
p.zza Leonardo da Vinci , 20133, Milano, Italy
phone: + 39- 02-23993328, fax: + 39-02-23993360, email: mariagabriella.signorini@polimi.it; federico.esposti@polimi.it
web: www.biomed.polimi.it

ABSTRACT

Nowadays a lot of methods for the estimation of Hurst's coefficient (H) in time series are available. Most of them, even if very effective, need some a priori information to be applied (in particular about the stationarity of the series). We analyzed eight up-to-date methods (working both in time and in frequency domain) at work with four kinds of synthetic time series (fBm, fGn, 1/f, FARIMA) in the range 0.1≤H≤0.9. We built graphs for each method evaluating the quality of the estimation, in terms of accuracy (bias) and precision (STD) of the deviation from the expected estimation value. Beginning from that, we formulated a procedure useful for a reliable estimation of H, using these existing methods, without any assumption on the stationarity of the time series. This procedure suggests to estimate, at first, the coefficient "alpha", spectral slope in a bi-logarithmic scale-estimator chart, next to the zero-frequency axis, of the unknown time series. Once estimated alpha, i.e. an indirect estimation of the stationarity of the series, the procedure recommends the best method for the estimation of H, depending on the stationarity value.

***Key words**: Hurst's Coefficient, Stationarity, Periodogram α coefficient*

1. INTRODUCTION

In last decades, complete dedication of non-linear signal processing was paid to the evaluation of skilled methods for the estimation of Hurst's coefficient in long memory fractal time series [1].
Nowadays the trend is to overcome classical approaches "Hurst way", e.g. the Rescaled Range Analysis, because of the strong estimation bias for series with H>0.7, and Autocorrelation methods, useful only for large and persistent time series.
The aim of this paper is to evaluate the effectiveness of the most used up-to-date tools for the estimation of the self-similarity coefficient in completely unknown time series. The point is that some *a priori* information is required for a proper use of most of the existing estimation methods. E.g., the relation between the interpolated slope coefficient α of the log-log plot scale-estimator of most up-to-date methods and the coefficient H depends from the series stationarity. Very often these information are not available. Our aim is to plan a procedure useful to properly estimate the value of Hurst's coefficient, using the existing methods, without any *a priori* knowledge about the stationarity of the signal.

2. MATERIALS AND METHODS

In this paper we considered eight different tools for the evaluation of Hurst's coefficient: the Method of the Aggregate Variance [2], the Method of the Modulus of the Aggregate series [2], the Higuchi method [3], the Dispersional Analysis (DA) [4], the Detrended Fluctuation Analysis (DFA) [5], the Periodogram method [2] [6], the Allan factor method and the Fano factor method [7].

The Aggregate Variance is a time domain method useful for non-stationary time series that obtains the multi-scale analysis with the aggregation of adjacent points and measures the similarity in terms of variance. The Modulus of the Aggregate series method is similar to the previous one but it uses modulus instead of variance. The Higuchi method is a time domain method useful for non-stationary series too, but performs the multi-scale analysis with the creation of sub-series, in following iterations, with points taken at different distances each other. In this case the similarity is described beginning from the partial sums of the original time series (derived from sub-series) and finding a normalized length (Higuchi length) of the sub-series. DA is the differential version of the Aggregate Variance method, useful for stationary series. DFA is the well-known estimator with detrend, working in the time domain. It is effective both with stationary and non stationary time series. The Periodogram is a frequency domain method, suited for stationary time series, that evaluates the slope of the spectrum (calculated by a Discrete Fourier Trasform) near the zero-frequency axis in a log-log plot. The value of this slope is correlated to the Hurst's coefficient by known relations. Finally Allan and Fano factors are frequency domain methods, very close each other, that indirectly evaluate the value of the slope mentioned for Periodogram without the calculation of the Fourier Trasform, used for stationary time series, too.

The reliability of these methods was tested by applying theme to synthetic time series with a known Hurst coefficient. These series belong to the four most used methods in literature: fractional Brownian motion (fBm), fractional

Gaussian noise (fGn), 1/f power law noise and fractional auto-regressive integrated moving-average (FARIMA) model [8]. These numeric series have different mathematical and statistical properties, so it was possible evaluating the effectiveness of the estimation methods, in very different conditions. The fBm is a generalization of the concept of Brownian motion, in which the innovation between a sample and the successive one is not necessarily un-correlated. In this way a non-stationary stochastic process is obtained. Directly derived from fBm, by discrete derivation, is the process called fGn that produce a stationary stochastic time series. The 1/f power law noise (later 1/f) is a synthetic non-stationary time series obtained by the anti-transformation of a synthetic spectrum (with uniformly distributed phase), built with a desired slope. FARIMA(0,H,0) (with Gaussian innovations), at last, is a sampling of a fGn process, built with a fractional integration of order d (with d = H + 1/2) i.e. a convolution with a filter h_d whose z-trasform is $h_d(z) = (1-z)^{-d}$ [9], of an ARMA(0,0) process.

The performance evaluation of the presented methods was made by applying each method to five realization of each synthetic time series. A "realization" is composed by nine numerical series of 50000 points, one for each value of H in $0.1 \le H \le 0.9$, with step 0.1. Were non-existent in literature, the required relations to estimate H beginning from the results of the different methods were investigated. For clarity, in Table 1, we list all the relationships between the value of slope obtained from the bi-logarithmic scale-estimator chart of each method and the Hurst's coefficient for each kind of synthetic time series analyzed. Once calculated all H values, we built confidence statistics (mean and STD) for estimations and we judged each method in terms of deviations from linearity and bias. Finally, for each series generation model and for each value of H, we evaluated which was (were) the best method(s). Figure 2 summarizes the results.

3. RESULTS

We obtained graphs for each method representing mean value and standard deviation for each kind of synthetic time series at each H value ($0.1 \le H \le 0.9$) (not shown here, see [10]). Beginning from these we built graphs representing the deviation of the estimation made by each method, for each kind of time series, for every $0.1 \le H \le 0.9$ and from the expected estimation value. By studying this last series of graphs we observed (see Figure 2):

fBm series: the estimators operating in time domain best fitted the expected values for fBm series, in the central region of the range $0<H<1$. Is detectable a polarization of estimations toward H=0.5. As expected, Periodogram performs unsatisfactory evaluations for non-stationary time series.
For the whole range $0.1 \le H \le 0.9$, we point out: for fBm with $H<0.3$ the best method is DFA; for $H>0.6$ Aggregate Variance, Modulus of the Aggregate series and Higuchi are suggested. In the range $0.3 \le H \le 0.6$ the performances of all the time domain methods are similar.

fGn series: in the whole range $0.1 \le H \le 0.9$ the performances of DFA, DA and Periodogram are similar. Fano and Allan factor are satisfactory only for H=0.1.
1/f series: for $0.3 \le H \le 0.8$ the most reliable methods are Modulus of the Aggregate series, Higuchi and DFA. The estimations performed by Periodogram are affected by a constant bias so its use is advisable for time series with boundary values of H (H near to 0^* and 1^-) where others methods appear as unreliable.
FARIMA series: no method had reported excellent performance for a meaningful range of variation of H.

As expectable, a first distinction is detectable: non-stationary time series are best analyzed by time domain methods and stationary series by frequency domain methods. DFA assures good quality performances for both stationary and non-stationary time series. This characteristic derives from the operation of detrending operated in DFA. Moreover DFA is characterized by a relation between the slope obtained from the log-log scale-estimator plot and the coefficient H which is identical for non-stationary and stationary real time series (i.e. fBm, fGn and 1/f). It is necessary here noting that the DFA directly estimates the spectral slope (usually called α, here P for Periodogram and DFA), as Periodogram, that have a continuous range ($-1<\alpha<3$) for both stationary and non stationary time series (For $-1<\alpha<1$ the series examined is stationary; for $1<\alpha<3$ the series is non-stationary and for $\alpha<-1$ or $\alpha>3$ the series is not a fractal time series [11] [12] [13]). Unfortunately Periodogram fails in non-stationary series H estimation because it is based on Fourier Trasform.
Beginning from the above remarks an analysis procedure useful to properly estimate the value of Hurst's coefficient without any *a priori* knowledge of the signal can be inferred. This procedure is:
First step: estimation of P (or α) of the unknown time series with DFA. Beginning from this first estimation of P, it's possible to value the "stationarity degree" of the series. The boundary value of P between stationarity and non-stationarity is $\alpha = 1$. We have decided to distinguish three different "stationarity zones" in α range. "High stationarity" for $\alpha<0.5$, "High non-stationarity" for $\alpha>1.5$ and a transition zone for $0.5 \le \alpha \le 1.5$.
Second step: for High stationary time series the best estimations are assured by DA or Periodogram; for High non-stationary series are recommended Modulus of the Aggregate series, Higuchi or DFA itself and for intermediate behaviors the value of H from DFA of step one is reliable.

4. DISCUSSION

The graphs obtained in this work are an useful handbook for the Hurst analysis of time series. Once estimated H for an unknown series, in fact, is possible to evaluate the quality of the estimation, in terms of accuracy (bias) and precision (Standard Deviation). In this stage, these remarks are valid for good-Signal-to-Noise-Ratio (SNR) time series.
The main target got in this paper is the formulation of the analyzing procedure for the blind estimation of Hurst's exponent for time series. The procedure, in step two, suggests for

high stationary and high non-stationary zones a couple of methods. It's probably opportune to apply both methods proposed and make the mean of the results. This opportunity should be evaluated examining the graphs presented above.
As explained before, the advantage of a blind method for the estimation of H is to avoid the choice between the relations useful for stationary time series and non-stationary time series (Fig.2). But why not simply use a stationarity index instead of DFA in step one? An obvious answer is because if the time series is highly non stationary or have a intermediate behavior the result of step one is already the H estimation. A less obvious answer is because in the range of variation of α the boundary line between stationary and non-stationary behavior is punctual. This is a problem because for some methods the relation between P (or α) and H is not the same for stationary and non stationary time series (for example for Periodogram). A little error during the estimation of α, i.e. during the evaluation of stationarity, could produce a big error in terms of H. So it's advisable, in our opinion, use a method that directly estimates α with a good linearity in the neighborhood of α =1.
The method proposed presents some criticality in presence of important discontinuity of the signal. For example with Lévy flight processes [14]. We experimented this condition with a biological signal (see **Example of application**): the fetal heart rate (FHR) signal. This process, for long registrations, can be assumed as a Lévy flight. Considering stretches without changes in the "fetal activity", however, FHR is assimilable to an fBm process (P≈1.5 "transition zone"). Analyzing the whole signal, was necessary to replace DFA with Periodogram. Only a frequency domain method, in fact, where is calculated the spectral slope in proximity of the zero-frequency axis, is insensitive to very high frequency contributes. We verified that the Hurst's exponent estimated for the whole time series by this "modified method" equals that estimated by the "normal one" for stretches without changes in the "fetal activity". This trick however, can be used only for time series with α≤1.5. In very high non-stationary series, as seen above, Periodogram method is useless.
Furthers developments of this work must be the analysis of the performances of methods varying the number of points of the series and, above all, the noise superimposed to it.
Build the graphs of deviation from the expected estimation value at different noises, will make simpler to comment the accuracy and the precision of the estimate in "real cases". Finally, in the number of the methods analyzed, must be included other up-to-date methods not still studied in this context; for example we refer to the Wavelet Trasform Modulus Maxima (WTMM) [9].

5. EXAMPLE OF APPLICATION

We tested the effectiveness of the procedure with the FHR series [15]. The fetal heart rate (FHR) is the recording of the electrical activity of the heart of the fetus. The FHR series is a fractal time series characterized by periodical "changing of activity" that, from the signal analysis point of view, assimilate FHR to a Lévy Flight process. We applied our procedure, with the shrewdness shown in discussion, i.e. using Periodogram in the First step, to evaluate the Hurst's coefficient of heart rate series of healthy fetuses (53 recordings) and of fetuses affected by intra-uterine grow retardation (IUGR) (46 recordings). IUGR regards particularly a retardation in the autonomous nervous system development. We first applied Periodogram and found a mean α value approximately equal to 1.6 for both populations. That means a border line value between transition zone and non-stationarity zone. This result is in accordance with the literature [16] [17]. So we decided, according to the "Second step", to analyze stretches without changes in the "fetal activity" for both groups with DFA and obtained that coefficient H is able to distinguish the two population (see Fig. 1).

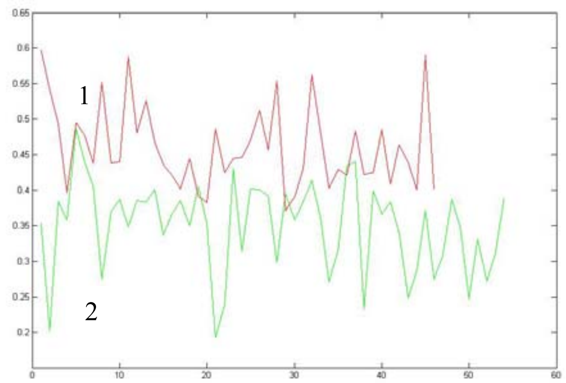


Fig. 1: H values estimated for healthy patients (line 2) and IUGR patients (line 1)

Method	Relation for fBm and 1/f	Relation for fGn and FARIMA
Aggregate Variance	H = P/2 + 1	/
Modulus of the Aggregate series	H = P + 1	/
Higuchi	H = P + 2	/
DFA	$H = P_{(2)} - 1$	$H_{fGn} = P_{(2)} - 1, H_{FA} = 2P_{(2)} - 1$*
DA	/	$H_{fGn} = P + 1$, H_{FA} *emp.**
Periodogram	H = (P – 1) / 2	H = (P + 1) / 2
Fano factor	/	$H_{fGn} = P/2$, $H_{FA} = P - 1$*
Allan factor	/	$H_{fGn} = P/2$, $H_{FA} = P - 1$*

* marks all the relation calculated in our work *emp*. Marks an empirical relation, fitted from data.

Tab. 1: List of the relationships between the value of slope obtained from the bi-logarithmic scale-estimator chart of each method (P) and the Hurst's coefficient (H) for each kind of synthetic time series analyzed.

This result derive from the autonomous nervous system development retardation. This lets up systemic controls and made the sum of central and peripheral heart rate controls less "chaotic".

6. CONCLUSIONS

In this work we obtained a whole of statistics to evaluate the quality of the estimations of the Hurst's exponent made by different estimation methods. The analysis of the performances was made for different kinds of synthetic time series (fBm, fGn, 1/f and FARIMA), for each value of H in the range $0.1 \leq H \leq 0.9$ with step 0.1, for eight different methods: Aggregate Variance, Modulus of the Aggregate series, Higuchi, Dispersional Analysis (DA), Detrended Fluctuation Analysis (DFA), Periodogram, Allan factor and Fano factor. We obtained statistics made by the mean and standard deviation of five estimations on different time series and we built graphs representing the deviations from the expected value, for each method, for each kind of time series and for each value of H. The graphs obtained are useful to evaluate the quality of the estimation, in terms of accuracy (bias) and precision (Standard Deviation). Beginning from these graphs we elaborated a procedure able to estimate the Hurst's coefficient without any knowledge about the stationarity of the series in exam. This procedure is made up from two steps:

First step. Estimation of P (or α) of the unknown time series with DFA. Beginning from P, it's possible to value the stationarity of the series. We have decided to distinguish three different "stationarity zones" in α range. "High stationarity" for $\alpha<0.5$, "High non-stationarity" for $\alpha>1.5$ and a transition zone for $0.5\leq \alpha\leq 1.5$.

Second step. Estimate of H as follows: for High stationary time series the best estimations are assured by DA or Periodogram; for High non-stationary series are recommended Modulus of the Aggregate series, Higuchi or DFA itself and for intermediate behaviors the value of H from DFA of step one is reliable.

We are now developing this work on other estimation methods (as WTMM) and at different conditions of SNR.

REFERENCES

[1] J. Beran, *Statistics for Long Memory Processes*, Chapman & Hall, N.Y., 1994

[2] M.S. Taqqu, V. Teverovsky, W. Willinger, *Estimators for long-range dependence: an empirical study,* "Fractals", 1996

[3] T. Higuchi, *Approach to an irregular time series on the basis of the fractal theory,* Physica D 31: 277-282, 1988

[4] J.B. Bassingthwaighte, G.M. Raymond, *Evaluation of the Dispersional Analysis Method for Fractal Time Series,* Annals of Biomedical Engineering, Vol. 23, 491-505, 1995

[5] C-K. Peng, S. Buldyrev, S. Halvin, *Mosaic organisation of DNA nucleotides*, Phys. Rev. E vol. 49: 1685-1689, 1994

[6] P. Flandrin, *On the spectrum of fractional Brownian motions*, IEEE transaction on information theory, vol. 35, N 1: 197-199, 1989 Jan

[7] R. Scharf, M. Meesmann et al., *General relation between variance-time curve and power spectral density for point processes exhibiting $1/f^{\beta}$ –fluctuations, with special reference to heart rate variability*, Biol.Cybern. 73: 255-263, 1995

[8] P.J. Brockwell, R.A. Davis, *Time Series: Theory and Methods*, Springer Verlag, N.Y., 2nd edition, 1991

[9] B. Audit, E. Bacry et al., *Wavelet-Based Estimators of Scaling Behaviour*, IEEE transaction on information theory, vol. 48, N 11: 2938-2954, 2002 Nov

[10] F. Esposti, *Valutazione di metodi di stima del coefficiente di autosomiglianza: applicazione a serie temporali sperimentali,* Master thesis in Bioengineering, Politecnico di Milano, 2005

[11] B. Pilgram and D.T. Kaplan, *A comparison of estimators for 1/f noise*, Physica D, vol. 114, pp. 108-122, 1995

[12] G. Magenes, M.G. Signorini, R. Sassi, *Automatic diagnosis of fetal heart rate: comparison of different methodological approaches*, Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE Volume 2, 25-28 Oct. 2001 Page(s):1604 - 1607 vol.2

[13] M.G. Signorini, R. Sassi, S. Cerutti, *Working on the Noltisalis database: measurement of nonlinear properties in heart rate variability signals*, Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE Volume 1, 25-28 Oct. 2001 Page(s):547 - 550 vol.1

[14] P. Lévy, *Théorie de l'Addition des Variables Aléatories*, Gauthier-Villars, Paris, 1937

[15] F. Esposti, M.G. Signorini, M. Ferrario, G. Magenes, *Self-similarity behavior characterization of Fetal Heart Rate signal in healthy and Intrauterine Grow Retardated fetuses*, submitted to IEEE EMBC New York, 2006

[16] T.J. Bigger., R.C. Steinam et al., *Power law behaviour of RR interval variability in healty middle-aged persons, patients with recent acute myocardial infarction, and patients with heart transplants*, Circulation, 93(12), 2142-2151, 1996

[17] M.G. Signorini, G. Magenes, S. Cerutti, D. Arduini, *Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings*, IEEE Trasactions on Biomedical Engineering, vol.50, no.3, 2003.

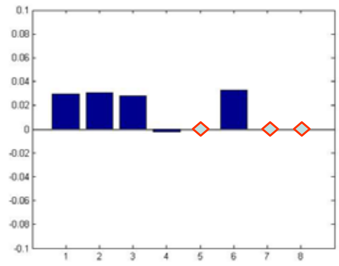
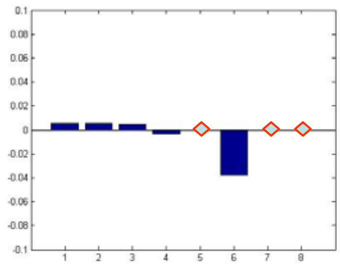
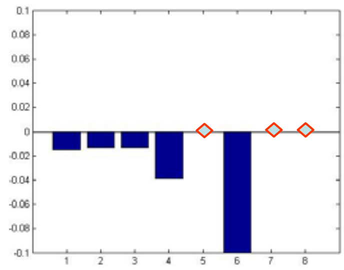
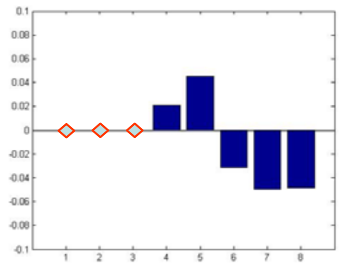
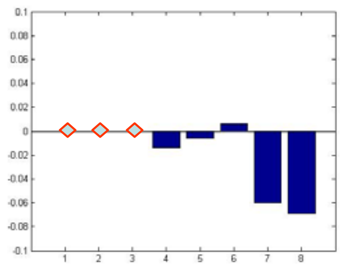
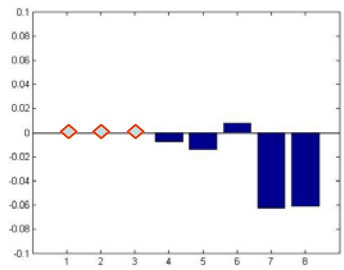
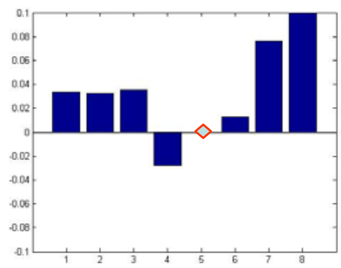
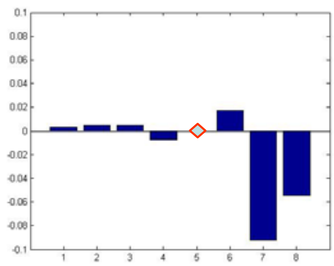
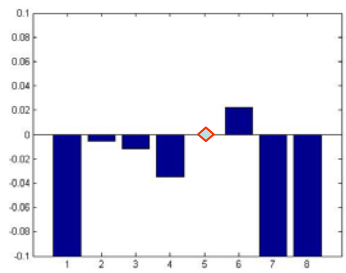
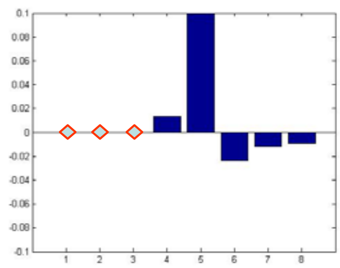
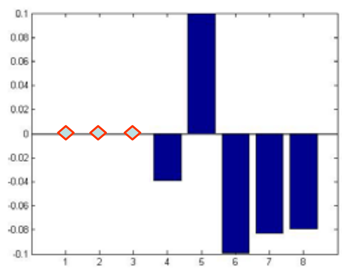
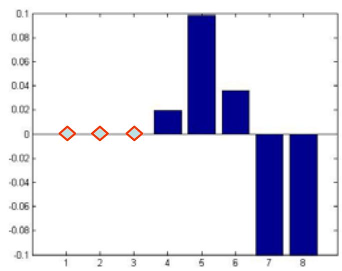
	H=0.2($\alpha_{fBm,1/f}$=1.4; $\alpha_{fGn,Farima}$= -0.6)	H=0.5 ($\alpha_{fBm,1/f}$= 2; $\alpha_{fGn,Farina}$ = 0)	H=0.8 ($\alpha_{fBm,1/f}$=2.6; $\alpha_{fGn,Farima}$= 0.6)
fBm			
fGn			
1/f			
Farima			

Fig. 2: Graphs representing, for each kind of time series, the deviation of the estimation, made by each method for H=0,2 H=0,5 and H=0,8, from the expected estimation value. 1=Aggregate Variance, 2=Modulus, 3=Higuchi, 4=DFA, 5=DA, 6=Periodogram, 7=Fano, 8=Allan.

◇ = Methods not used for the synthetic series in exam.